# Classification de phonèmes à base d'apprentissage profond - Projection dans un contexte de parole dégradée

Sondes ABDERRAZEK[1]
Corinne FREDOUILLE[1]
Alain GHIO[2]
Muriel LALAIN[2]
Christine MEUNIER[2]
Virginie WOISARD[3]

[1]LIA, Avignon Université
[2]Aix-Marseille Univ, LPL, CNRS, Aix-en-Provence
[3]UT2J, Octogone-Lordat, Toulouse Université & Toulouse Hospital

Les mesures perceptives restent la méthode la plus courante pour évaluer les troubles de parole en pratique clinique. Leur subjectivité en terme de reproductibilité ainsi que leur manque d'interprétation (ex. la localisation des altérations) motivent le développement d'outils d'évaluation objectifs.

La littérature propose de nombreux travaux sur l'évaluation objective de la sévérité/intelligibilité, la détection automatique de troubles de la parole ou leur classification [1][2][3].

Nous proposons ici d'étudier un système basé sur des réseaux de neurones profonds pour une tâche de classification automatique de phonèmes [4]. Ce travail est la première étape d'un projet à long terme, qui vise à déterminer les unités linguistiques contribuant au maintien ou à la perte de l'intelligibilité chez des patients atteints de troubles de la parole. L'intérêt porté aux systèmes neuronaux repose sur les performances démontrées en traitement de la parole, mais, également, sur la volonté de voir ces systèmes neuronaux autrement qu'une simple « boîte noire ». Il devient important aujourd'hui de comprendre et expliquer le fonctionnement de ces systèmes et leurs

décisions -- travaux sur l'interprétabilité et l'explicabilité [5][6][7] -- éléments essentiels pour des applications médicales notamment.

Ici, une architecture neuronale de type « Convolutional Neural Network » (CNN) est entrainée sur de la parole lue (corpus BREF [8]) pour une tâche de classification de phonèmes. Le modèle est ensuite testé sur de la parole pathologique issue de patients traités pour un cancer de la tête ou du cou vs des témoins (tâche de lecture du corpus C2SI [9]). L'objectif de cette première étude est d'analyser la réponse du modèle CNN aux troubles de la parole afin d'étudier ultérieurement son efficacité à fournir des connaissances pertinentes en termes de perte d'intelligibilité.

L'évaluation du CNN repose sur le calcul d'un coefficient de corrélation (Pearson - r) entre les scores de classification obtenus par le CNN à partir des productions des sujets sains et patients et des mesures perceptives sur la qualité de la parole fournies par un jury d'experts.

Le tableau 1 montre que la sévérité et l'altération phonémique corrèlent le mieux avec les sorties du CNN. Cette observation est cohérente du fait que celles-ci s'approchent le plus de la tâche de classification visée en terme d'altération acoustique globale ET locale perçue des unités phonémiques. L'intelligibilité, quant à elle, est associée à une corrélation moindre, probablement liée à une surestimation de la mesure par les experts due au phénomène d'habituation au texte lu utilisé.

*Tableau 1*. Corrélation entre les différentes mesures perceptives et les performances de classification du CNN

| **Mesure perceptive** | Intelligibilité | Sévérité | Altération phonémique |
|---|---|---|---|
| **r** | 0.78 | 0.91 | –0.88 |

## Références bibliographiques


[1] T. B. Ijitona, J. J. Soraghan, A. Lowit, G. Di-Caterina, and H. Yue, *"Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification,"* in IET 3rd International Conference on Intelligent Signal Processing (ISP 2017), 2017, pp. 1–6.

[2] B. Vachhani, C. Bhat, B. Das, and S. K. Kopparapu, *"Deep autoencoder based speech features for improved dysarthric speech recognition,"* Interspeech 2017, 2017, pp. 1854–1858.

[3] L. Bin, M. C. Kelley, D. Aalto, and B. V. Tucker, *"Automatic speech intelligibility scoring of head and neck cancer patients with deep neural networks,"* in International Congress of Phonetic Sciences (ICPHs'19), Melbourne, Australia, 2019.

[4] S. Abderrazek, C. Fredouille, A. Ghio, M. Lalain, C. Meunier, V. Woisard, "Towards Interpreting Deep Learning Models to Understand Loss of Speech Intelligibility in Speech Disorders Step 1 : CNN model-based phone classification", Interspeech'20, Shanghai, China, 2020.

[5] T. Pellegrini and S. Mouysset, *"Inferring phonemic classes from CNN activation maps using clustering techniques,"* in PROCEEDINGS OF INTERSPEECH'16, San Francisco, US, 2016

[6] T. Nagamine, M. L. Seltzer, and N. Mesgarani, *"Exploring how deep neural networks form phonemic categories,"* in PROCEEDINGS OF INTERSPEECH'15, Dresden, Germany, 2015

[7] H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specte, Lalana Kagal, *"Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning,"* in 5th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2018)

[8] L. F. Lamel, J. L. Gauvain, and M. Eskenazi, *"BREF, a large vocabulary spoken corpus for french,"* in Proceedings of European Conference on Speech Communication and Technology (Eurospeech'91), Genoa, Italy, 1991

[9] C. Astesano, M. Balaguer, J. Farinas, C. Fredouille, P. Gaillard,A. Ghio, L. Giusti, I. Laaridh, M. Lalain, B. Lepage, J. Mauclair,O. Nocaudie, J. Pinquier, O. Pont, G. Pouchoulin, P. Michele,D. Robert, E. Sicard, and V. Woisard, *"Carcinologic Speech Severity Index Project: A Database of Speech Disorders Productions to Assess Quality of Life Related to Speech After Cancer,",* in LANGUAGE RESOURCES AND EVALUATION CONFERENCE (LREC), Miyazak , Japon, may 2018